MODELING AND ANALYSIS OF MASS CASUALTY TRIAGE SYSTEMS

By

BAILEY WHITMAN BONJOUR

Bachelor of Science in Industrial Engineering and Management

Oklahoma State University

Stillwater, Oklahoma

May, 2017

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 2020

MODELING AND ANALYSIS OF MASS CASUALTY

TRIAGE SYSTEMS

Thesis Approved:

Dr. Manjunath Kamath

Thesis Adviser

Dr. Camille Deyong

Dr. Chaoyue Zhao

Name: BAILEY WHITMAN BONJOUR

Date of Degree: MAY, 2020

Title of Study: MODELING AND ANALYSIS OF MASS CASUALTY TRIAGE SYSTEMS

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: In the aftermath of a mass casualty incident, a large number of patients are likely to arrive at a hospital for medical care. The large patient demand often overwhelms the capacity of the medical resources available in an event called a patient surge, in which medical triage is often utilized. This study develops analytical models based on queuing theory that can be used as the basis for a tool to determine what staffing and other hospital resources are needed during a patient surge. Additionally, this study presents a simulation model that can also be utilized alongside the analytical models to analyze details that are not captured in the analytical approach. These models allow different patient volume and makeup scenarios to be evaluated so that the resources needed can be estimated. The models and codes developed in this study could be paired with a decision support system that hospital administrators and planners could use to develop contingency plans for mass casualty incidents with a variety of patient volumes and makeup. Finally, this study also made a small contribution to the queueing body of knowledge by extending results available for a Markovian multi-server priority queue to yield simple and reasonably accurate approximations for the general multi-server priority queue.

Table of Contents

Chapter Page

LIST OF TABLES ... vi
LIST OF FIGURES ... vii
CHAPTER I ... 1
1.1 Background ... 1
1.2 Introduction to Triage ... 3
1.3 Introduction to Queuing Theory ... 5
1.4 Brief Literature Review ... 5
1.5 Thesis Outline ... 6
CHAPTER II ... 8
2.1 Disaster Management and Operations Research ... 8
2.2 Queuing, Simulation, and Hospital Operations ... 9
2.3 Surge Capacity ... 10
2.4 Reducing Door to Doctor Time ... 12
2.5 Reducing Pre-Triage Time ... 12
CHAPTER III ... 16
3.1 Problem Statement ... 16
3.2 Goals ... 16
3.3 Objectives and Tasks ... 17
3.4 Scope and Limitations ... 19
3.5 Research Contribution ... 20
CHAPTER IV ... 22
4.1 Triage Process Outline ... 22
4.2 Physical Arrangement ... 25
4.3 Modeling Approach ... 26
CHAPTER V ... 30
5.1 Two-Priority Triage System Overview ... 30
5.2 Modeling Initial Triage ... 31
5.3 Modeling the Operating Rooms ... 32

Chapter Page

5.4 Input to the Treatment Areas 37
5.5 Modeling the Treatment Areas 39
5.6 Summary of Analytical Models Developed 41
CHAPTER VI 43
6.1 Main Triage System 43
6.2 Treatment Area Systems 45
CHAPTER VII 48
7.1 Initial Triage Numerical Case 48
7.2 Operating Room Numerical Case 49
7.3 Treatment Areas Numerical Case 57
CHAPTER VIII 60
8.1 Summary of Work 60
8.2 Contributions 61
8.3 Future Work 62
BIBLIOGRAPHY 63
APPENDIX I 67

LIST OF TABLES

Table Page

Table 1- Healthcare Providers Needed for Triage Classifications ... 24
Table 2 – Operation Time Breakdown ... 49
Table 3 – List of Combinations Tested ... 51
Table 4 - High Priority Results ... 58
Table 5 - Low Priority Results ... 59

LIST OF FIGURES

Figure Page

Figure 1- Physical Triage Arrangement 26
Figure 2 – Triage Queuing Network 31
Figure 3 – Treatment Area Network 39
Figure 4 – Main Triage Simulation 44
Figure 5 – Treatment Area Simulation 46
Figure 6 - High Priority Wait Time Results with 90% Utilization and Exponential Service Time Distribution 53
Figure 7 - Low Priority Wait Time Results with 90% Utilization and Exponential Service Time Distribution 53
Figure 8 - High Priority Wait Time Results with 90% Utilization and Triangular Service Time Distribution 54
Figure 9 - Low Priority Wait Time Results with 90% Utilization and Triangular Service Time Distribution 54
Figure 10 - High Priority Wait Time Results with 97% Utilization and Exponential Service Time Distribution 55
Figure 11 - Low Priority Wait Time Results with 97% Utilization and Exponential Service Time Distribution 55
Figure 12 - High Priority Wait Time Results with 97% Utilization and Triangular Service Time Distribution 56
Figure 13 - Low Priority Wait Time Results with 97% Utilization and Triangular Service Time Distribution 56

CHAPTER I

INTRODUCTION

1.1 Background

In the aftermath of a large-scale disaster, hospitals are put under immense pressure. A large surge of injured patients requiring immediate surgical treatment floods the hospital system. As the number of patients arriving over a short period of time often exceeds the capacity of the hospital, healthcare professionals are faced with more patients than they can quickly treat, leaving patients to wait in a queue until medical professionals become available.

This scenario is referred to as a mass casualty incident (MCI), defined as any incident in which medical resources are overwhelmed by the number and severity of casualties. Patient surges often occur after MCIs such as building collapses, mass-transportation accidents, and natural disasters. Despite the sharp increase in patient demand following an MCI, it is imperative that patients are seen and cared for promptly.

Hospital operations after an MCI face many obstacles, depending on the type of disaster. During an epidemic, hospitals face a surge in demand while experiencing a decrease in available healthcare workers as medical professionals become ill and are unable to care for others. In the aftermath of a natural disaster, such as a hurricane or an earthquake, power and water shortages can be expected and roads may be blocked or destroyed, slowing the arrival of supplies or volunteer healthcare professionals. After an MCI, it is imperative that emergency procedures are in place to ensure that patients that were in the hospital system before the disaster and patients in the hospital system as a result of the disaster are both cared for in a timely manner.

Although there are certainly cases where patient outcomes may not be significantly affected by an extended wait before care is received, there are numerous cases where an immediate surgical treatment is one of the most important factors in a successful patient outcome. Even initially non-life-threatening injuries have the potential to become life threatening after an extended period of time due to the threat of infection.

In the hours following an MCI, it is crucial that hospitals have a plan in place to care for the surge of incoming patients. This is referred to as surge capacity, or the ability expand beyond normal services to manage a sudden influx of patients. The World Health Organization states that all hospitals should have a preparedness plan for patient surge to prevent chaos during mass casualty management and to treat and save as many patients as possible (World Health Organization). The volume and makeup of patients vary depending on the MCI type. Additionally, each hospital has different resources available, ranging from available personnel to operating rooms. A decision support tool has the potential to aid hospitals while creating their preparedness plans to ensure that the available resources specific to each hospital are used effectively, resulting in a higher level of patient care.

1.2 Introduction to Triage

When a large surge of patients arrives at a hospital unexpectedly, triage is the most commonly used method of managing patients with limited resources. Patients are placed into different categories depending on the severity of their injuries and the state of the patient. Patients are then taken to the appropriate area for further monitoring as they wait for admission to the operating room. Typically, patients with the most severe injuries but that are likely to survive are the highest priority.

The use of triage originated in military operations. Although triage principles were likely used previously, the first records of triage were described by Baron Dominique Larre (1776-1842), a surgeon in the Napoleonic Army. Larre's methods "included initial treatment and triage of the wounded and triage of the wounded on the battlefield before transport by horse-drawn "ambulances" to hospitals located in the rear." Before the use of triage, wounded soldiers were left until the conclusion of the battle and then collected, often by order of rank (Kennedy, Aghababian, Gans, & Lewis, 1996).

Since its inception as a military operation, triage is now used almost universally in the aftermath of MCIs. Although the challenges hospitals face during a surge event vary depending on what type of event has occurred, a common issue in the Emergency Department is reducing the period of time between a patient's arrival and the time the patient is seen by the triage healthcare provider in charge of evaluating and classifying patients. This metric is often referred to as Pre-Triage Time (PTT), and is crucial because it allows medical professionals to determine which patients need to be admitted into an operating room most urgently. Failure to quickly perform this initial screen could result in a severely injured patient's condition further declining as they wait to be evaluated.

Another important Emergency Department (ED) metric is the Door to Doctor (D2D) time, or the time between a patient's arrival to the ED and the patient seeing a physician, physician assistant, or nurse practitioner for their needed treatment. If a patient is in critical condition, D2D time is incredibly important to ensure the patient's condition does not further deteriorate. Long D2D times are correlated with a higher number of sick patients that leave without being treated (J. Cochran & Burdick, 2011). The median D2D time is around 30 minutes. The shortest D2D time of 12 minutes was seen in patients considered to have an immediate need to be seen ("Multivariable testing cuts door-to-doc times by 24%," 2007). Just as the D2D time is important in an emergency department, a quick D2D time after an MCI is critical to ensure patients receive care in a timely manner.

In an Emergency Department (ED), most facilities use the time of initial triage and registration as the time of arrival (Houston, Sanchez, Fischer, Volz, & Wolfe, 2015). However, this ignores the time between patient arrival time and triage time, which is referred to as pre-triage time (PTT), as mentioned above. In normal ED operating conditions, studies have shown this time to reach as high as 98.6 minutes (Betz, Stempien, Trivedi, & Bryce, 2017) to 105 minutes (Houston et al., 2015), with a median PTT of 11 minutes, which is approximately 30% of the mean national door-to-doctor time. Unsurprisingly, longer pre-triage wait times are more prevalent in emergency department peak hours. Building on this correlation, it stands to reason that if the pre-triage time is a concern during surges that are typical to normal ED operating conditions, it could be an even more significant concern during a surge following an MCI.

1.3 Introduction to Queuing Theory

After an MCI, the demand (patients needing to be treated by a healthcare professional) outnumbers the capacity (healthcare professionals' ability to provide care). In this case, there are patients waiting to be treated by healthcare personnel that form a queue, as the healthcare personnel providing service cannot outpace the demand of patients flowing into the hospital. Queuing theory involves a rigorous mathematical study of queues or "waiting lines" that form when the rate of arrival (in this case, influx of patients into a hospital) temporarily exceeds the rate of service (patients that can be treated and discharged).

Queuing theory became a popular field of study after World War II as the result of work by Agner Erlang as a way to model incoming telephone calls. Since its origin, it has been applied in many areas, including banks, computer systems, communication systems, manufacturing systems, and hospitals. It is typically used to estimate waiting times and determine how many service providers are necessary to achieve reasonable waiting times and an acceptable level of resource utilization.

1.4 Brief Literature Review

Although queuing theory originated shortly after World War II, its applications in hospitals were not widespread until later. One of the earliest investigations looked into the queuing process in hospital outpatient departments and the patient's waiting time (Bailey, 1952). The same researcher later established a capacity threshold where the service rate equals the arrival rate (Bailey, 1954). After these investigations, research regarding queuing applications in healthcare seemed to slow until the 1970s. Since then, several advances have been made in using queuing theory to estimate bed requirements, determine staffing needs, and limit wait times.

Although queuing theory has been applied to several areas of healthcare, there is still relatively very little research over queuing applications in hospitals relating to surge capacity. In disaster situations, when the demand is significantly greater than a hospital's capacity to provide care, having the ability to expand service is invaluable. Abujudeh, Vuong, and Baker investigated the use of portable X-ray examination procedures in an emergency room (Abujudeh, Vuong, & Baker, 2005). De Bruin, Koole, and Visser used queuing theory to analyze the cause of bottlenecks in the emergency department (de Bruin, Koole, & Visser, 2005).

If not treated promptly, victims of a disaster could face severe, life-changing or life-threatening consequences. Because of the severe consequences of belated treatment and the potential volume of patients affected by a disaster, it is imperative that further research take place to investigate methods of treating patients quickly and effectively in the aftermath of a disaster.

1.5 Thesis Outline

Chapter 2 contains a literature review over research relevant to meeting the surge in demand following a disaster. Following this, chapter 3 lists the research objectives of this project. Chapter 4 then details the triage modeling approach. This includes the current triage process, physical arrangement, and the queuing approach. Chapter 5 explains the analytical models developed in this study. Chapter 6 describes the simulation model developed in Simio and the tests used to validate it. Chapter 7 presents the numerical cases tested as a part of this study. In this chapter, results from all three sections of the triage model are examined. Finally, Chapter 8 presents a summary of the work completed as a part of this study, further work, and the contribution of this study.

CHAPTER II

LITERATURE REVIEW

2.1 Disaster Management and Operations Research

There have been numerous studies conducted to mitigate the effects of a large-scale disaster. These papers fall under disaster management, which is the organization and management of resources and responsibilities for dealing with humanitarian aspects of emergencies in order to lessen the impact of disasters ("About Disaster Management - IFRC"). There are four phases of disaster management: mitigation, preparedness, response, and recovery. ("Phases of Disaster: Disaster Preparedness and Economic Recovery,") The mitigation phase encompasses efforts that attempt to reduce vulnerability to disasters, such as enforcing building codes or constructing levees to protect a city. Preparedness focuses on the impact that a disaster would have on a community and includes education and training, as well as emergency planning. Response addresses the immediate effects of a disaster and includes triage and meeting humanitarian needs such as food and shelter, and cleanup. The recovery period is the restoration of a community to normalcy.

Operations research is frequently applied to disaster management problems, as it provides a quantitative foundation for decision-making. Disaster management applications of operations research are numerous and typically fall under the preparedness and response phases of disaster management. Some examples include Sung and Lee's work on using column generation to optimally allocate emergency resources after an MCI (Sung & Lee, 2016) and Fawcett and Oliveira's work on casualty treatment after earthquake disasters (Fawcett & Oliveira, 2000).

2.2 Queuing, Simulation, and Hospital Operations

Queuing theory and simulation have proven to be effective tools in improving hospital operations. Several studies have used queuing theory and simulation to determine the bed capacity of a hospital. Cochran and Bharti used queuing theory to analyze the utilization of beds within different areas of a hospital and then re-allocate the beds to fit the utilization patterns of the hospital (Cochran & Bharti, 2006). Cochran also researched hospital Emergency Department capacity needs using queuing theory in a subsequent paper (Cochran & Roche, 2009). Later, Pinto et al. used queuing theory and regression analysis to identify the minimum number of hospital beds that would meet the demand needs of a hospital (Pinto, De Campos, Perpetuo, & Ribeiro, 2015). Takagi, Kanai, and Misue also used queuing networks to analyze the flow of obstetric patients in a hospital, which could then be used for capacity planning of hospital wards in the future (Takagi, Kanai, & Misue, 2017).

Queuing theory and simulation have also been used in hospital applications to evaluate different methods of patient routing. Connelly and Blair explored the use of discrete event simulation (DES) to analyze the movement of staff between emergency department areas to compare the benefit of an alternate patient flow pattern (Connelly & Bair, 2004). Bish et al. evaluated the usage of patient segmentation using queuing theory, discovering that the split-flow method allowed healthcare professionals to see more patients (Bish, McCormick, & Otegbeye, 2016). Similarly, Bonalumi et al. used queuing theory to analyze the use of a ‘Super Track’ to see low-acuity patients at an inner-city hospital (Bonalumi et al., 2017).

Queuing theory and simulation have been used to improve hospital operations in other capacities as well. For example, Bahadori et al. used queuing theory and a simulation model to evaluate different server usage methods (Bahadori, Mohammadnejhad, Ravangard, & Teymourzadeh, 2014).

2.3 Surge Capacity

At the intersection of disaster management and hospital operations lies the concept of surge capacity, defined as the ability to manage a sudden, unexpected increase in patient volume that would otherwise severely challenge or exceed the present capacity of the facility (Hick, Barbera, & Kelen, 2009). In ‘Refining Surge Capacity: Conventional, Contingency, and Crisis Capacity’, Hick et al. describe four key interdependent factors that contribute to effective surge response: system, staff, space, and supplies. The authors proposed three subsets of overall surge capacity: conventional, contingency, and crisis. These three subsets correspond to different levels of surge, with conventional capacity consistent with daily practices, contingency capacity having little impact on usual patient care, and crisis capacity not consistent with usual standards of care but providing sufficient care in a disaster setting.

Although papers dealing with extraordinary surge remain relatively rare compared to the research conducted on daily surge events and ED overcrowding, several papers have addressed patient surges after a mass casualty incident. In 2005, Sacco et al. conducted one of the first attempts to analytically model resource-constrained patient prioritization, modeled as a classic resource allocation problem (Sacco et al., 2005). Ouyong et al. used a simulation model of a Lubbock hospital to evaluate the impact of a surge event on the hospital, identifying a scenario that would cause the model to crash (Ouyang, Patvivatsiri, & Montes, 2006). Later, Muhammet and Fuat also used discrete event simulation (DES) to identify that when the percentage of patients in critical condition exceeds 20%, more staff is needed (Muhammet & Ali Fuat, 2015).

Narrowing into the combination of queuing theory and surge capacity, several recent studies have used queuing theory to propose or evaluate methods of dealing with large quantities of patients following an MCI. Gong and Batta identified a queue-length cutoff model for a two-priority queue to treat as many patients as possible during a surge event (Gong & Batta, 2006). Cohen et al. used a two-stage tandem queuing system with flexible servers to evaluate surgeon allocation; surgeons are often the scarcest resource after an MCI (Cohen, Mandelbaum, & Zychlinski, 2013). Adalja et al. focused on absorbing the citywide patient surge following Hurricane Sandy using queuing theory, identifying that the long-term patient surge that followed the closure of many hospitals affected by Hurricane Sandy was more taxing on hospitals than the acute patient surge immediately following the hurricane (Adalja et al., 2014). Lodree, Altay, and Cook used queuing theory paired with optimization to evaluate the best staff assignment policies after an MCI, considering different patient prioritization levels and servers arriving sporadically, as would be the case following an MCI (Lodree, Altay, & Cook, 2017).

2.4 Reducing Door to Doctor Time

Significant research has been conducted on the reduction of door-to-doctor times, as D2D time is highly correlated to patient satisfaction and hospitals are graded on their D2D times. Research has been conducted to make the ED triage process ‘leaner’, reducing the number of steps patients and staff must go through before a physician sees a patient. In one hospital, these changes led to a D2D reduction from 67 minutes to 18 minutes ("Slash door-to-doc time, boost patient approval," 2011). One hospital that meets its goal of a D2D time of 31 minutes in 98.3% of patients cite good communication and constant awareness of waiting patients as keys to maintaining D2D times ("Busy ED keeps promise of 'door to doc' in 31 minutes," 2008). While D2D time is a common metric in hospital emergency departments, it is also a critical consideration after an MCI.

In a triage situation after an MCI, D2D time can have different meanings due to the number of physicians involved in the triage process. There are physicians utilized at the initial triage area, treatment areas, and at the operating room. For the purpose of this paper, D2D time will refer to the time between a patient’s arrival at the triage area and the patient’s admission into an operating room.

2.5 Reducing Pre-Triage Time

Although considerable research has been done on reducing time between the initial assessment of a patient and the treatment of a patient under surge conditions, very little literature has been identified that addresses the time between the arrival of a patient to a hospital and the initial assessment of that patient, which is referred to as pre-triage time (PTT). Only two papers have been identified that investigate the concept of pre-triage time (Betz et al., 2017; Houston et al., 2015). Both of these papers investigated the time between patient arrival and initial triage in an emergency department under normal conditions. Both papers also noted the lack of data on pre-triage time, as the time of arrival is usually noted as the time of initial triage screening in an emergency department setting, which ignores any patient wait between time of arrival and time of initial triage. Additionally, both papers noted that PTT increased when the emergency department experienced a surge.

However, a parallel can be drawn between PTT and the wait passengers experience before going through security screening at airports. In both situations, the number of people waiting typically significantly outnumbers the number of people acting as servers. Similarly, people are waiting in long queues before going through a screening or assessment process in both situations. Congestion grows as the number of waiting people increasingly outnumbers the number of servers. However, the consequences of prolonged wait times greatly differ between the two scenarios. A prolonged wait in a security queue will result in a missed flight in the worst-case scenario. In an emergency triage situation, prolonged wait times can result in greater bodily damage than the initial injury, or even death.

After the events of September 11, 2001, airport procedures in the United States have significantly evolved, resulting in a more thorough screening procedure that has increased average wait times before screening. The passenger screening process seeks to achieve two objectives: security and customer service, which is primarily achieved through timely wait times before screening (Gkritza, Niemeier, & Mannering, 2006). In 2016, airports, airlines, and the TSA scrambled to identify ways to reduce long security screening wait times after facing harsh criticism of wait times that reached as high as two hours during peak times ("TSA Boosts Efforts to Cut Waits at Airports -- WSJ," 2016). There is a considerable amount of literature focusing on reducing wait times at security screening. In 2016, Dorton and Liu investigated the impact of baggage volume and alarm rate on a security checkpoint's performance, finding that baggage volume had a moderate effect on cycle time and no effect on throughput (Dorton & Liu, 2016). Later, Song and Zhuang analyzed optimal screening policies, looking to balance security and congestion concerns (Song & Zhuang, 2017).

In initial triage and security screenings, the objective is the same: to screen a large amount of people in as little time as possible. Although there has been no literature discovered on improving the efficiency of initial triage screening, the existing literature on improving the airport screening process suggests that queuing principles could be applied to initial triage screening to prevent excessive wait times for initial triage screening.

In the triage setting, the quality of care patients receive while waiting for admission to an operating room is an additional concern not present in the security screening application. However, just as queuing theory principles can be applied to security screenings, they can be applied to triage to ensure that there is sufficient staffing so that patients receive the level of care needed for their condition.

The literature review shows that applications of operations research and queuing theory have been successfully completed in areas similar to triage. These principles have the potential to decrease waiting times and increase the quality of care a patient receives in triage after a mass casualty incident. By applying these principles to the triage setting, this study has developed tools that can be used to determine what resources are needed to decrease wait times and increase the quality of care patients receive. In doing so, this study contributes to the body of literature on applying queuing theory to determine resources needed in a triage setting.

CHAPTER III

RESEARCH STATEMENT, GOALS, AND OBJECTIVES

3.1 Problem Statement

After an MCI, the ability to quickly respond and treat patients is invaluable. In many instances, the time between injury and treatment could mean the difference between life and death for a patient. In a post-MCI scenario, the ability to appropriately treat patients rests on the implementation of an effective triage screening queuing system. An effective triage system must maintain a reasonable waiting time between a patient's arrival and the initial triage screening, a reasonable waiting time before admission to the operating room, and provide sufficient care during the patient's wait between initial screening and admission to the operating room. Efficient and sufficient staffing of the triage queuing system has the potential to make a significant positive impact on the wellbeing of the patient.

3.2 Goals

There is need for more research in developing models to determine resource needs either in a post-MCI or surge scenario to maintain a reasonable wait time before surgical treatment and a reasonable level of care while waiting for admission to the operating room.

This study developed two analytical models that can be used to analyze the resources necessary to meet patient needs under a range of patient makeup and volume scenarios following an MCI. One model focuses on wait time before initial triage screening and wait time before admission to the operating room. The other model focuses on the level of care in treatment areas. These models could then be used in a decision support system that emergency responders and healthcare professionals could use to determine the appropriate parameter levels needed in a specific scenario to achieve an appropriate triage waiting time and level of care for patients. In addition, the study also developed a simulation model in Simio to not only validate the analytical models, but also for potential use alongside the analytical models to support the decision-making process.

3.3 Objectives and Tasks

To achieve the research goals, the following research objectives and tasks were completed:

1. Published work in relevant areas was identified and reviewed.

   A thorough literature review was completed that focused on applications of queuing theory that are relevant to triage. Current triage methods and techniques were also explored in detail, as well as the complex problems that occur when dealing with patient surges and large-scale triage. Tasks associated with this objective included the following.

   a. Current triage methods and relevant queuing literature were reviewed
   b. Information regarding challenges and considerations present during patient surges was gathered and evaluated

c. Data was gathered on triage waiting times and staffing levels deemed medically safe

d. Relevant data was gathered on volume and makeup of patients after different types of MCIs

e. Current methods of triage were reviewed, including physical arrangement

2. Analytical models of the triage system were developed and tested

Developing analytical models of the triage process was the main contribution of this study. To develop the analytical models, the details of patient priority had to be captured correctly in order to reflect the impact the priority system has on waiting time of patients. Additionally, the care provided in the treatment areas had to be modelled. Tasks associated with this objective include the following.

a. An analytical model for a general muti-server priority queue was developed

b. The analytical model was validated by comparing its results to the output of a Simio model

c. An analytical model of the treatment areas was developed through a linkage with the output of the priority queue model.

d. Numerical experimentation of models was completed

3. A simulation model in Simio of the triage process was developed.

The simulation model served as a way to validate the main analytical model. Additionally, the simulation model has the ability to capture some details that the analytical model cannot and could play a complementary role. Completed tasks associated with this objective include the following.

    a. An effective way of modeling the priority queue in Simio was identified
    b. A model of the treatment areas that was tied to the main triage system was developed
    c. The simulation was numerically validated using exact results available for the special Markovian case of the priority queue model.
4. The analytical models were coded in Excel for future use in a decision support system

   A decision support system will allow first responders and hospital administrators to investigate the use of different staffing arrangements for triage and create plans for specific MCIs. When the codes are coupled with a user-friendly interface and the simulation model, the resulting decision support system could be valuable to hospital contingency planning.

3.4 Scope and Limitations

The focus of this project was to build analytical models appropriate for use in a what-if mode for hospital administrators and planners to identify resources needed to meet patient needs after an MCI. Both hospitals and MCIs vary greatly in terms of available staff, number of injuries, and makeup of injury types. Because of this, models that can be used to experiment with different parameters and analyze what resource levels will meet patient needs in a variety of situations is an obvious choice of tool.

The key metrics analyzed by the models are the mean of the wait before initial triage screening (PTT), mean of time between patient arrival and admission into an operating room (D2D), and the level of care patients receive as they wait for admission to an operating room.

The scope of the project is limited to the time between the patient's arrival at the triage area of a hospital (typically outside the main entrance of the hospital) and the patient's exit from an operating room.

Transportation to the triage area and evacuation are not included in the scope of this project. Additionally, identifying an optimal parameter setting or method is not an objective. Instead, the focus of the project is to identify how parameters can and should be shifted in order to obtain the level of care in the triage holding areas, wait before initial triage, and wait before admission to operating rooms deemed ideal by medical professionals.

3.5 Research Contribution

This project explored the parameters associated with triage and their impact on important aspects of patient care. The contribution of this study includes the development of a method for modeling the triage system, which has components that do not fit the typical queuing framework. Specifically, the treatment area of the triage system is atypical because it serves as a queue in which services are performed. Additionally, this study developed approximations for a general multi-server priority queue used for estimating the waiting time of patients for the operating room.

When used in a decision support system, the modeling of the triage system through both the analytical models and simulation will allow healthcare professionals and emergency planners to analyze how parameters affect key metrics that impact patient health as they prepare emergency plans for mass casualty events.

CHAPTER IV

MODELING APPROACH

4.1 Triage Process Outline

After an MCI, victims are transported from the scene of the event to a hospital. In normal hospital operations, when the number of patients needing treatment does not overwhelm hospital resources, patients are taken into the hospital for triage screenings and treatment. However, in a large patient surge such as one following an MCI, the triage area is usually set up outside the entrance of the hospital.

As patients arrive at the triage area, they are screened and sorted into categories depending on the severity of their injuries. Even patients arriving by ambulance that have already been evaluated should be re-triaged upon arrival.

There are several different triage methods medical professionals use when dealing with severe patient surge. This project assumes the use of SALT guidelines for mass casualty triage. SALT stands for "Sort, Assess, Life-saving interventions, and Treatment". This guideline was developed by an interdisciplinary committee formed by the CDC and comprised of the AMA, the American College of Surgeons, the American College of

Emergency Physicians, and the National Association of EMTs, as well as several other relevant organizations (Ugarte). SALT categories are as follows:

- Immediate: Requires attention within minutes to two hours upon arrival to avoid death or major disability
- Delayed: Wounded in need of surgery but whose condition permits treatment delay without unduly endangering life, limb, or eyesight
- Minimal: Have relatively minor injuries and can either effectively care for themselves or require only minimal care
- Expectant: Have injuries that overwhelm current medical resources. This category is re-evaluated frequently as resources become available.

Although four patient categories are described in the SALT guidelines, only two categories were considered in this study: immediate and delayed. These categories are referred to as high priority and low priority in the mathematical and simulation models. Initially, the decision to limit the categories examined to two was due to the shortcomings of the initial model used to reflect the priority queue. However, after a more comprehensive model was discovered, extended, and subsequently utilized in this study, modeling all four categories is now possible and easily implementable.

The triage officer, usually the most skilled surgeon or the medical professional with the most triage experience, is responsible for evaluating and categorizing each patient. Under some circumstances, the triage officer may also provide extremely quick life-saving procedures, such as unblocking an airway. When the number of patients waiting to be seen is extremely large, more than one triage officer may be utilized. In some situations, more than one treatment area per triage category may also be necessary (*Emergency War Surgery*, 2004).

For security purposes and to minimize chaos in the triage area, the flow into and through the triage area is primarily one-directional, with the exception of litter-bearers that carry patients to designated areas.

After patients have been placed in their initial categories, they continue to be monitored by medical professionals while waiting for their operations. Suggested staffing levels for immediate and delayed patient care are included in Table 1 (Cotter, 2006). Minimal patients typically do not receive medical attention until the other patients have been treated and expectant patients receive only basic care for comfort.

Table 1- Healthcare Providers Needed for Triage Classifications

Normal Operations	
Immediate Care (High Priority)	1 Advanced Life Support (ALS) provider and 1 Basic Life Support (BLS) provider per patient
Delayed Care (Low Priority)	1 BLS provider per patient and 1 ALS provider per 3 patients
Major Emergency Medical Operations	
Immediate Care (High Priority)	1 ALS provider and 1 BLS provider per 2 patients
Delayed Care (Low Priority)	1 BLS provider per 3 patients and one ALS provider per 5 patients
Disaster-Level Operations	
Immediate Care (High Priority)	1 ALS provider per 3 patients and 1 BLS provider per 5 patients
Delayed Care (Low Priority)	1 BLS provider per 5 patients and 1 ALS provider per 10 patients

Advanced Life Support providers complete advanced life-support measures such as giving medicine via injection. Basic Life Support providers do not provide any care that breaks the skin of the patient.

Treatment areas for all patient categories are designed to be easily expandable. This allows the physical space the patients occupy to accommodate the necessary number of patients. Because treatment areas serve as the holding area before a patient is admitted into an operating room, the treatment area and staffing assigned to the treatment area must be able to accommodate the entire queue of patients waiting for the operating room(s). While patients are in the treatment area, care must be provided periodically by healthcare personnel.

Triage is an inherently dynamic process. If the patient's condition changes significantly while in a holding area, the patient will be re-triaged into the correct category. Refined triage often takes place within a category, with the medical professional over an area determining which patients have first priority to be admitted into surgery.

In triage situations, the operating room is consistently the bottleneck. Patients in the immediate category are admitted into the operating room first, followed by delayed patients. There can be any number of operating rooms present in a triage system. The modeling of the operating room(s) was an important piece of this study. The modeling of the operating room(s) had to capture the details of the patient priority or category in order to accurately show the average number of patients that waited in both treatment areas as well as the average amount of time spent there, depending on their priority level.

4.2 Physical Arrangement

The triage model shown in Figure 1 focuses only on the immediate and delayed categories. The number of triage officers, number of patient beds, and number of life support providers are parameters that can be adjusted to control the level of patient care and wait time before initial triage evaluation.

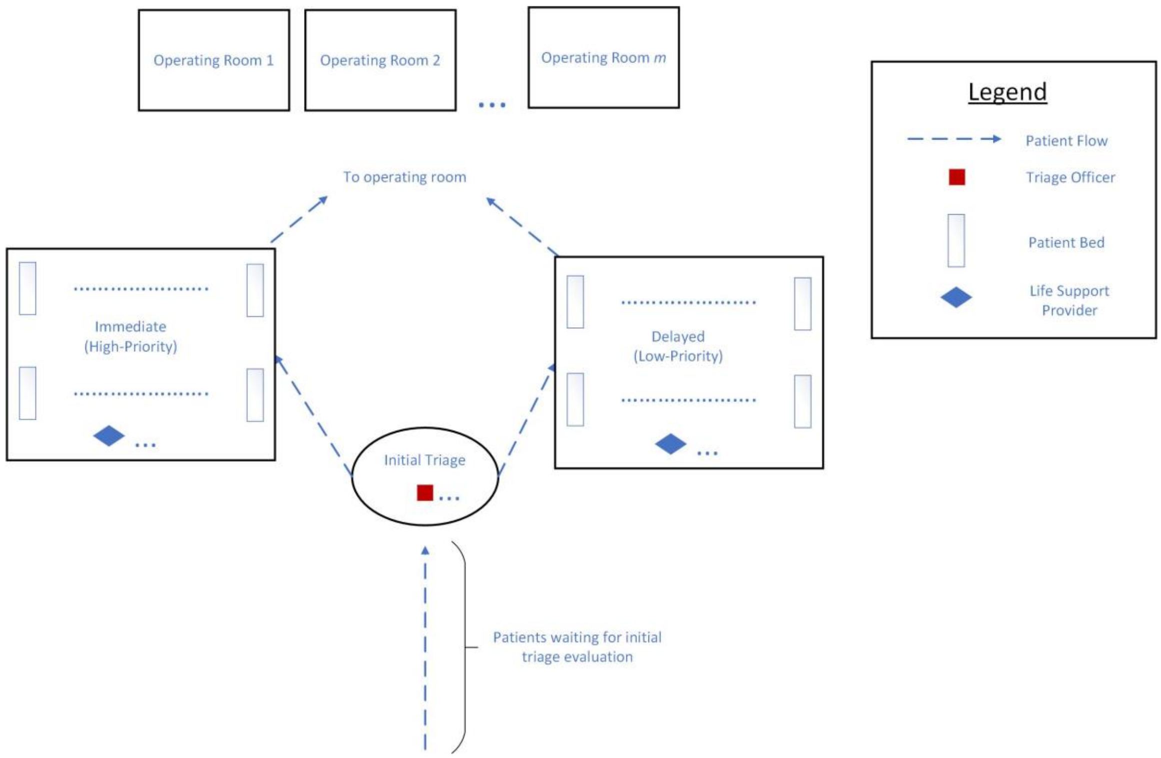

Figure 1- Physical Triage Arrangement

4.3 Modeling Approach

A queuing network is a clear choice to model and analyze the performance of the triage system. Queuing networks are often used to analyze the performance of complex systems such as computers, communication networks, and production shops (W Whitt, 1983). The triage queuing network model is comprised of nodes and arcs. The nodes represent the areas where patients receive care of any type. The arcs represent the flow of patients. Patients enter the network when they arrive at the triage area and begin waiting for an initial triage screening. Patients leave the network when they leave the operating room. For this project, Whitt's two moment approach and arrival and departure mechanisms will be used as described in his 1983 paper, *The Queuing Network Analyzer*.

In a typical queuing network, if all servers at a node were busy, patients would join a queue and wait until a server is free. When the server becomes free, the service of the next patient in line is carried out without interruption. This is true for the initial triage screening node in the triage system. This node may have one or more servers, which are triage officers in this context. After service is complete at this node, patients are sent to a treatment area node. The percentage of patients sent to each node is likely determined by the type of MCI that caused the patient surge.

Although the triage screening node acts as a traditional node, the treatment areas for each patient type do not follow the same patterns as typical service nodes. Instead of patients waiting for their individual turn with a server, as in a typical node, in treatment areas, patients do not wait before being admitted into the treatment area. As the number of patients in a treatment area increases, the level of care patients receive in treatment areas may decline unless more servers are added to care for patients. Treatment area nodes were modeled in a two different ways for this study. In the main triage system, treatment areas were modeled as queues for patients waiting for admission to the operating room(s). Using this approach, the average number of patients in the treatment areas and average time a patient spends in the treatment areas was calculated. However, as patients wait in the queue for admission to the operating room, the service provided by medical professionals monitoring patients and preventing further deterioration in the patient's condition must also be modeled.

To model the care patients receive in the treatment areas, a secondary model was utilized that used the time patients spend in the treatment areas. Using this information, the number of medical professionals needed to maintain the necessary level of patient care could be calculated by analyzing the delay experienced by patients in the treatment area in receiving the care they require from the treatment area staff.

Patients move between nodes. In the triage model, this may be because their service is complete (in the screening node) or because an operating room has become open. The patient moves between nodes until the patient exits the system (through the operating room in this case). The queuing network model for a triage system is considered open because patients arrive from and return to the outside of the system.

Following Whitt's approach, key elements in the general approach for modeling a queuing network are as follows:

1. Parameters are used to characterize the flows and service at nodes. Because these parameters can easily be changed, the model can be applicable to many different situations.
2. Approximations for multiple server queues are based on the partial information provided by the parameters that characterize the arrival process and service-times at each node.
3. Calculus for transforming parameters represents the basic network operations of merging, splitting, and departure
4. A synthesis algorithm solves the system of traffic rate and variability equations resulting from the basic calculus applied to the network.

The general approach of a using a queuing network is to represent all arrival processes and service time distributions using a few parameters. This will allow different scenarios to easily be modeled once the basic arrangement is in place, allowing analysis to be conducted to determine what staffing and bed levels are needed in the aftermath of a MCI with varying characteristics.

CHAPTER V

ANALYTICAL MODEL OF THE TRIAGE SYSTEM

5.1 Two-Priority Triage System Overview

In this modeling approach, only the immediate and delayed patient categories are considered and are denoted as high priority and low priority, respectively.

Notation

$\lambda = Total\ Arrival\ Rate\ of\ Patients$

$c_a^2 = Squared\ Coefficient\ of\ Variation\ of$ *Inter-arrival time*

$\tau_T = Average\ Service\ Time\ of\ Initial\ Triage$

$c_{s,T}^2 = Squared\ Coefficient\ of\ Variation\ of Initial\ Triage\ Service\ Time$

$m_T = Number\ of\ Physicians\ at\ Initial\ Triage$

$p = Percentage\ of\ Low\ Priority\ Patients$

$\tau_O = Average\ Time\ for\ Surgery$

$c_{s,O}^2 = Squared\ Coefficient\ of\ Variation\ of\ Surgery\ Time$

$m_O = Number\ of\ Operating\ Rooms$

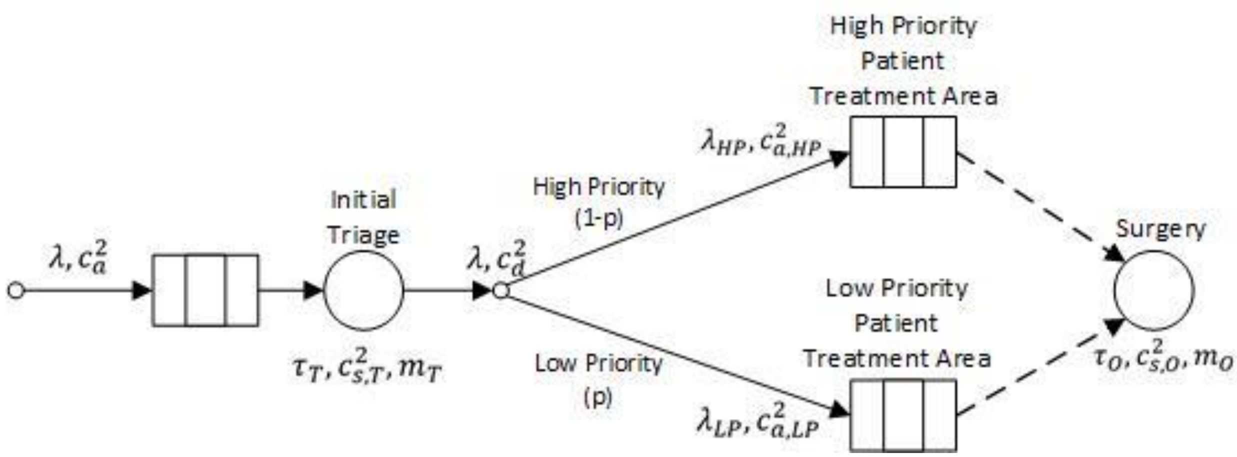

Figure 2 – Triage Queuing Network

5.2 Modeling Initial Triage

Patients arrive to the hospital according to an arrival process with rate λ and squared coefficient of variation c_a^2. These patients then form a FIFO queue for the initial triage server. Initial triage has an average service time, a squared coefficient of variation, and specified number of triage officers that evaluate patients and assign them to a high or low priority level.

The formula shown below was used to approximate the waiting time incoming patients would experience at the initial triage node. The approximation is the simple version of the approximations suggested by Whitt (1993). A more complicated version of the approximation used later in the priority queue model could also be employed here.

$$\approx \left(\frac{c_a^2 + c_s^2}{2}\right) EW(M/M/m)$$

In the above formula, EW(*M/M/*m) is the average waiting time in queue for the corresponding multi-server queue with Poisson arrivals and exponential service times (see Appendix I for details).

Patients then exit initial triage. The squared coefficient of variation of the departure process can be derived from Whitt's approximation in *The Queuing Network Analyzer*, equations 38 and 39 (Whitt, 1983).

$$c_d^2 = 1 + \left(1 - \left(\frac{\lambda\tau_T}{m_T}\right)^2\right)(c_a^2 - 1) + \frac{(\lambda\tau_T/m_T)^2}{\sqrt{m_T}}(c_{s,T}^2 - 1)$$

5.3 Modeling the Operating Rooms

It is assumed that the surgical operating room has identical service times for patients of both priority levels. The mean service time and squared coefficient of variation for both priority levels are τ_O and $c_{s,O}^2$ respectively. The number of operating rooms (and required personnel teams) is represented by m_O.

At the initial triage, patients are classified as low priority with probability p, and as high priority with probability (1 – p). Patients enter the high priority treatment area after exiting initial triage with an arrival rate of

$$\lambda_{HP} = \lambda(1 - p)$$

and a squared coefficient of variation of inter-arrival time

$$c_{a,HP}^2 = (1 - p)c_d^2 + p$$

Patients enter the low priority treatment area after exiting initial triage at an arrival rate of

$$\lambda_{LP} = \lambda p$$

and squared coefficient of variation of inter-arrival time

$$c_{a,LP}^2 = pc_d^2 + 1 - p$$

High priority patients experience the triage system almost as if only high priority patients exist within the system. Within the queue, they have priority over low priority patients. The only time that high priority patients experience delay due to low priority patients is if a low priority patient is in surgery when a high priority patient arrives and all operating rooms are busy. In that case, the high priority patient must wait until the low priority patient's surgery is complete before they can begin their surgery. In other words, there is no preemption of service.

However, low priority patients are impacted heavily by the presence of high priority patients within the triage system. The operating room admits low priority patients into surgery only if there are no high priority patients within the triage system.

Because patients are treated differently in the system based on their priority levels, their waiting time in the triage system must be calculated using a priority-discipline queuing model, where the queue discipline is based on a priority system (Hillier and Lieberman, 2004). There are two types of priority discipline queueing models: non-preemptive and preemptive. In a non-preemptive queuing model, if a low priority customer is being served while a high priority customer arrives, the low priority customer's service will continue until completion, at which time the high priority customer's service will begin. In a preemptive system, if a low priority customer is being served while a high priority customer arrives, the low priority customer's service will be interrupted so that the high priority customer's service can begin immediately upon their arrival. In our case, a low priority patient's operation cannot be paused in the event of a high priority patient's arrival. Therefore, the triage system can be modeled as a non-preemptive priority discipline queueing system.

Hillier and Liberman (1984) provide exact results for a non-preemptive, muti-server priority queue with Poisson arrivals and exponential service times. Using their notation, W_k, the expected waiting time in the system for a patient of priority class k, can be calculated as:

$$W_k = \frac{1}{A * B_{k-1} * B_k} + \frac{1}{\mu}$$

where $A = s!\left(\frac{s\mu - \lambda}{r^s}\right) \sum_{j=0}^{s-1} \frac{r^j}{j!} + s\mu$

$B_0 = 1$

$B_k = 1 - \frac{\sum_{i=1}^{k} \lambda_i}{s\mu}$

and $s = number\ of\ servers$

$\mu = mean\ service\ rate\ per\ busy\ server = \frac{1}{\tau_O}$

$\lambda_i = mean\ arrival\ rate\ for\ priority\ class\ i$

$\lambda = \sum_{i=1}^{N} \lambda_i$

$r = \frac{\lambda}{\mu}$

N = number of priority classes

For this triage system application, high priority is class 1 and low priority is class 2. The expected time a class k patient will wait in the queue is equal to:

$$EW_k = \frac{1}{A * B_{k-1} * B_k}$$

The above approach models the expected wait time of high and low priority patients assuming that both the service and inter-arrival time distributions are exponential. When inter-arrival times and service times do not follow an exponential distribution, Whitt's "Approximation for the GI/G/m Queue" is used to approximate the estimated waiting time of patients (Whitt, 1993). Our rationale for applying Whitt's correction factor for non-exponential cases is based on the following observations.

The service time distributions are identical for both high priority and low priority patients, and the priority discipline is non-preemptive. Because of this, the average number in queue and the average waiting time in queue for an "aggregate" patient should be the same as in an equivalent multi-server queue with an arrival rate equal to the total arrival rate of all patients.

In Whitt's (1993) paper, the equation used to approximate the estimated wait time for general arrival and/or service parameters using the results from the M/M/m queue is as follows:

$$EW(\rho, c_a^2, c_s^2, m) \approx \phi(\rho, c_a^2, c_s^2, m)\left(\frac{c_a^2 + c_s^2}{2}\right) EW(M/M/m)$$

Where:

$$\phi(\rho, c_a^2, c_s^2, m) = \begin{cases} \left(\frac{4(c_a^2 - c_s^2)}{4c_a^2 - 3c_s^2}\right)\phi_1(m,\rho) + \left(\frac{c_s^2}{4c_a^2 - 3c_s^2}\right)\Psi\left(\frac{c_a^2 + c_s^2}{2}, m, \rho\right), c_a^2 \geq c_s^2 \\ \left(\frac{c_s^2 - c_a^2}{2c_a^2 + 2c_s^2}\right)\phi_3(m,\rho) + \left(\frac{c_s^2 + 3c_a^2}{2c_a^2 + 2c_s^2}\right)\Psi\left(\frac{c_a^2 + c_s^2}{2}, m, \rho\right), c_a^2 \leq c_s^2 \end{cases}$$

And

$$\Psi(c^2, m, \rho) = \begin{cases} 1, & c^2 \geq 1 \\ \phi_4(m,\rho)^{2(1-c^2)}, & 0 \leq c^2 \leq 1 \end{cases}$$

$$\phi_4(m,\rho) = min\left\{1, \frac{\phi_1(m,\rho) + \phi_3(m,\rho)}{2}\right\}$$

$$\phi_3(m,\rho) = \phi_2(m,\rho)^{\frac{-2(1-\rho)}{3\rho}}$$

$$\phi_2(m,\rho) = 1 - 4\gamma(m,p)$$

$$\phi_1 = 1 + \gamma(m,\rho)$$

$$\gamma(m,\rho) = min\left\{0.24, \frac{(1-\rho)(m-1)((4+5m)^{\frac{1}{2}} - 2)}{16m\rho}\right\}$$

After calculating $\phi(\rho, c_a^2, c_s^2, m)$, the EW_k calculated using the two-priority system can be utilized as shown below:

$$EW_H \approx \phi(\rho_O, c_{a,O}^2, c_{s,O}^2, m_O)\left(\frac{c_a^2 + c_s^2}{2}\right) EW_1$$

and

$$EW_L \approx \phi\left(\rho_O, c_{a,O}^2, c_{s,O}^2, m_O\right)\left(\frac{c_a^2 + c_s^2}{2}\right)EW_2$$

Where

$$\rho_O = \lambda\tau_O/m_O$$

$$c_{a,O}^2 = c_d^2$$

c_d^2 is the squared coefficient of variation of inter-departure time at the initial triage node.

5.4 Input to the Treatment Areas

Patients exiting initial triage are transported to the high and low priority treatment areas as they wait for admission into the operating room. Treatment areas act as the queue for the operating room. As patients wait for surgery in these treatment areas, they are closely monitored by medical personnel and given medications and other interventions as needed.

The arrival rate to the treatment areas is the same as the arrival rate to the operating room(s), as the treatment areas serve as a holding area for patients before they are admitted into the operating room. Therefore, as patients leave the initial triage node and arrive at the treatment areas, the arrival parameters are as before:

Patients enter the high priority treatment area after exiting initial triage at an arrival rate of

$$\lambda_{HP} = \lambda(1-p)$$

and a squared coefficient of variation of inter-arrival time

$$c_{a,HP}^2 = (1-p)c_d^2 + p$$

Patients enter the low priority treatment area after exiting initial triage at an arrival rate of

$$\lambda_{LP} = \lambda p$$

and squared coefficient of variation of inter-arrival time

$$c_{a,LP}^2 = pc_d^2 + 1 - p$$

Where c_d^2 is the squared coefficient of variation of inter-departure time at the initial triage node, and p is the percentage of patients that will be classified as low priority.

Because the treatment areas serve as the queue for the operating room(s), the average time a patient spends in the queue for the operating rooms is equal to the total time the patient is in the treatment area. Therefore, the average waiting time for the operating room serve as input for the treatment area models. The average number of patients in each treatment area can be calculated as follows.

$$L_{q,HP} = \lambda_{HP} EW_{HP}$$

$$L_{q,LP} = \lambda_{LP} EW_{LP}$$

The above would give us the average number of beds needed in each of the treatment areas. Next, an analytical model to evaluate the level of care that can be expected from different staffing requirements is presented. The average delay before a patient receives care could be used as a surrogate measure for the level of care. This model can then be used to determine the staffing required for adequate coverage in the treatment areas

5.5 Modeling the Treatment Areas

Once the patients arrive at the treatment areas, they are checked and tended to periodically. We model each treatment area as a two-node open queueing network as shown in Figure 3. The Check node represents the care that a patient receives from a healthcare worker periodically, and the Stable node models the time between checks where the patient is resting and requires no care. The “Exit” from the Stable node represents a patient leaving the treatment area to enter an operating room.

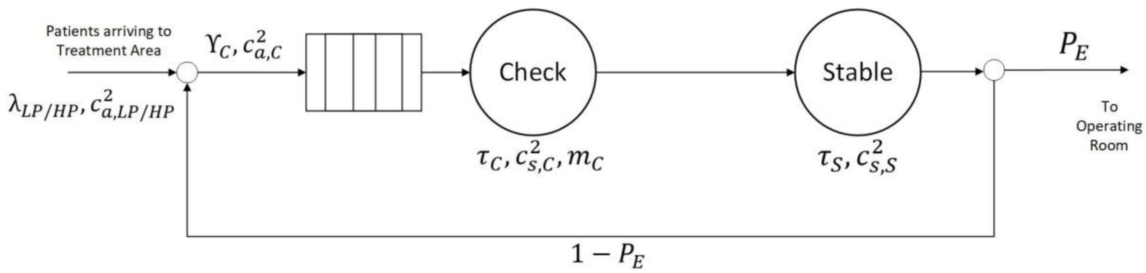

Figure 3 – Treatment Area Network

The patient’s exit from a treatment area is determined by the patient’s admission into the operating room. In this model, the probability of the patient exiting a treatment area is depicted as P_E. This probability is not known, but we know that its value should be set in such a way that the average time a patient spends in the 2-node network is equal to the average time spent by a patient waiting in queue for an operating room. Hence, the average time is system for the high-priority (low-priority) treatment area model should be equal to EW_{HP} (EW_{LP}) obtained from the priority queue model. This gives us a way to device a search procedure to find P_E. Of course, P_E values will be different for the high-priority and low-priority treatment area network models, Once, we have P_E, then the average delay before check, EW_c can be computed. The complete set of equations for the treatment area network model are given below.

In this network, Expected Total Time in Treatment Area was calculated as:

Notation

τ_C = Average Check Time

$c_{s,C}^2$ = Squared Coefficient of Variation of Check Time

m_C = Number of Medical Staff at Check

τ_S = Average Stable or Resting Time

P_E = Probability of Exit

γ_C = Total Arrival Rate to Check Node

$c_{a,C}^2$ = Squared Coefficient of Variation of Combined Arrival Processes to Check

$$Total\ Time = (EW_C + \tau_C + \tau_S) * \frac{1}{P_E}$$

$$EW_C \cong EW(M(\Upsilon_C), M\left(\frac{1}{\tau_C}\right), m_C * \frac{c_{a,C}^2 + c_{s,C}^2}{2}$$

$$c_{a,C}^2 = \frac{(1-P_E)^2\left\{\rho_C^2 + \frac{\rho_C^2}{\sqrt{m_C}}\left(c_{s,C}^2 - 1\right)\right\} + P_E\{1 - P_E + c_{a,HP}^2\}}{[1-(1-P_E)^2(1-\rho_C^2)]}$$

$$\rho_C = \frac{\Upsilon_C \tau_C}{m_C}$$

$$\gamma_C = \frac{\lambda_{(HP\ or\ LP)}}{P_E}$$

The treatment area network model was implemented in Excel and using the Solver capability in Excel, the $P_E$value was calculated for a specific set of parameter values. Once P_E is identified, the average wait time for the check node could be calculated. If the estimated wait for the Check node is excessive, it can be concluded that patients cannot be checked as frequently as desired and additional healthcare workers are necessary. The search procedure can be repeated with a new value for m_c till a satisfactory level of wait time is achieved.

The average number of patients in a treatment area,

$$L_{q,HP} = \lambda_{HP} EW_{HP}$$

$$L_{q,LP} = \lambda_{LP} EW_{LP}$$

calculated using Little's Law, is indicative of the average number of beds needed in the treatment area. The above results are important metrics because they reflect the number of healthcare workers and beds needed in the treatment areas.

5.6 Summary of Analytical Models Developed

Three main areas were modeled as part of this study: initial triage, operating rooms, and treatment areas. Each of these areas build off each other; the departure parameters from the initial triage node had to be calculated before the operating room calculations could begin. The waiting times from the operating room priority model were fed into the treatment area calculations in order to examine the level of care received under different staffing parameters. These models were verified using the estimated results from a Simio model. The models developed in this section are powerful tools that can be used to plan necessary staffing and bed space needed for patient care after an MCI, and their use has the potential to help hospitals obtain valuable information so they can be prepared with a plan before tragedy strikes.

CHAPTER VI

SIMIO MODEL DEVELOPMENTS

The simulation model was built using Simio software (Simio Student Edition Version 11, 2019). The simulation models two interdependent systems: the main triage system and the treatment areas. Although the simulation was built to validate the analytical models, the simulation model can be used as a planning tool on its own to fine tune the solutions developed using the analytical models. The simulation model has the ability to capture more information and model more details compared to the analytical models.

6.1 Main Triage System

The main treatment system consists of the initial triage server, the two treatment areas where high priority and low priority patients wait admission to the operating room, and the operating rooms where patients are eventually admitted. The queues for each operating room correspond to the treatment areas where patients receive care as needed before admission to the operating room. In the picture below, Server1 models initial triage, where a triage officer completes a quick scan of incoming patients to determine if the patient should be categorized as high priority or low priority. After initial triage, the patient moves to either HighPriority or LowPriority, both of which are treatment areas, before moving to the operating room and eventually exiting the system.

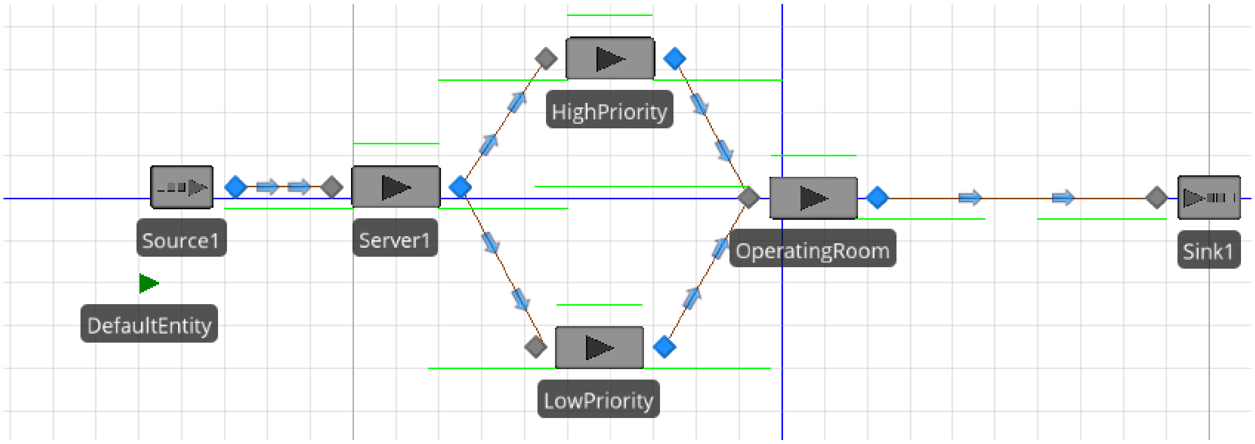

Figure 4 – Main Triage Simulation

The initial triage server acts on a first come first served priority. The rate of arriving patients can be modified in the MainSource properties. The service rate of the initial triage officer and the number of triage officers at the initial triage server can be easily modified in the InitialTriage server properties.

After being examined at the initial triage area, patients go to either the high priority or low priority treatment areas while they wait for admission to the operating room. The probability of a patient being sent to each treatment area is determined by the link weight of paths leading to each treatment area. The probability of these links can be easily edited to reflect the changes to the system that occur when the mix of high and low priority patients changes.

Admission to the operating room is priority based with no preemption. To model one operating room serving all patient categories, the capacity is set to 1. The ranking rule is set to highest priority first, which ensures that high priority patients are seen before low priority patients. Differing numbers of operating rooms can easily be modeled by changing the capacity of the server. This can show the impact of having different numbers of operating rooms ready to accept patients in a triage situation.

To quickly validate the Simio model, Little's Law was used to ensure that the average number of patients shown to be in the treatment areas made sense given the other parameters.

$$Little's\ Law: L = \lambda W$$

Where L= the average number of patients in the system, λ is the arrival rate, and W is the average waiting time. When checked across several parameter combinations, Little's Law held true as expected, validating the data retrieval methods used.

After the simulation was completed and partially validated using Little's Law, the simulation results were compared to exact analytical results available for the special exponential case for validation.

6.2 Treatment Area Systems

The treatment area systems are used to model the care that patients receive while waiting for admission to the operating room. The treatment area model for each triage category consists of a category-specific Source and two servers labelled Stable and Check.

When a patient in the main system enters the path to a treatment area, the source for that triage category is triggered to create a new entity. That entity then circulates through the Stable and Check servers. These servers model the frequency with which patients receive care. For instance, if a patient in a specific category must receive care every ten minutes, the processing time for the Stable server is ten minutes before the patients moves to the Check server. The Check server service time can be edited easily to model different processes that may need to take place while monitoring patients. The Stable server has infinite capacity, and the Check capacity is reflective of the number of care-givers present in the treatment area that is being modelled.

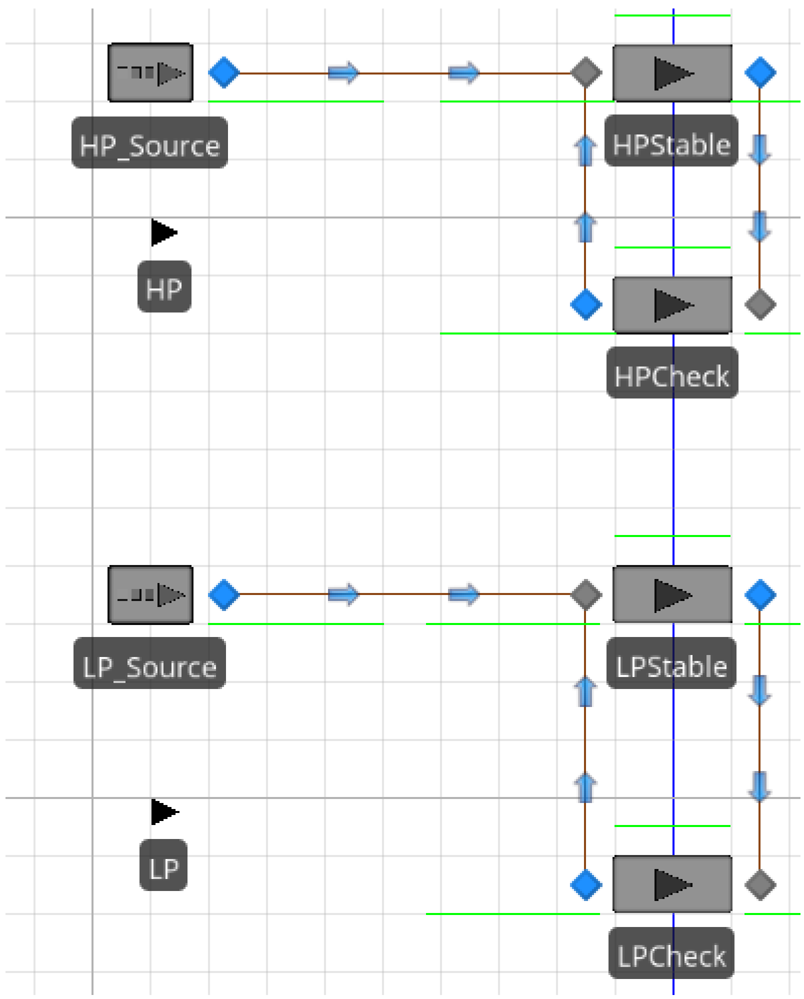

Figure 5 – Treatment Area Simulation

When a patient in the main triage system exits the operating room, an add-on process is triggered to remove one patient from the treatment area corresponding to the priority level of the patient that exited the main triage system. This ensures that the number of patients in each treatment area is equal to the number of each patient type waiting for the operating room in the main triage system.

CHAPTER VII

NUMERICAL EXPERIMENTATION

The main focus of the numerical experimentation was to validate the priority queueing model for non-exponential cases and demonstrate the efficacy of the treatment area models in producing the desired results. For analytical queueing models, the utilization per server has to be strictly less than one to yield valid results. To capture the heavy load on the hospital system during a surge, we tested the priority queue model at fairly high utilizations and chose 90% and 97% for our numerical experiments.

7.1 Initial Triage Numerical Case

To model the initial triage process, it was assumed that the initial triage service time followed a triangular distribution with a mean of 2 minutes, a maximum of 3 minutes, and a minimum of 1 minute. Two arrival rates were considered: $\lambda = 2.2383$ per hour and $\lambda = 2.07$ per hour. These arrival rates corresponded to 97% and 90% operating room utilization rates respectively. We assumed a Poisson arrival process, and c_a^2 was set equal to 1. Using the triangular(1,2,3) distribution to model the initial triage service time yielded a c_s^2 of 0.0833. A low variability in the initial triage service times is reasonable and even desirable as the main purpose is to quickly categorize patients and send them to the appropriate treatment areas

After completing the analytical calculations, it became clear that the wait time at initial triage was negligible (much less than a minute), due to the arrival rates that were selected to keep the utilization of the operating rooms under one. However, if the operating room capacity is significantly increased, the arrival rates could be much higher and at some point the triage wait times may become a problem, even to the point of necessitating an additional triage officer.

7.2 Operating Room Numerical Case

To identify the parameters that should be used in the priority queue model, the first step was to identify typical service times in surgical triage situations. "Duration and Predictors of Emergency Surgical Operations – Basis for Medical Management of Mass Casualty Incidents" stated that the average time in surgery during a triage was 130 minutes (Huber-Wagner et al., 2009). The text then went on to describe the breakdown of surgery times as shown below.

Table 2 – Operation Time Breakdown

Percentage of Patients	**Time in Operating Room**
54.1%	137 minutes
26.3%	110 minutes
11.5%	136 minutes
5%	91 minutes
3.1%	142 minutes

From this data, two potential service time cases were established. In the first, service time would be distributed exponentially, with the mean = 130. This would result in a c_s^2 equal to 1.

In the second case, a triangular distribution was used to model the data. Using the triangular distribution, the lower limit (a) =110, the upper limit (b) =142, and the mode (c) = 137. Using these parameters, the mean of the triangular distribution equals 130 minutes, which is consistent with the data stated above. Using these parameters, c_s^2 is equal to 0.0029

After establishing the service time parameter, the arrival rate was established. Because triage is only necessary when the number of patients entering a hospital nears or exceeds the capacity of the hospital to care for those patients, only high utilization cases needed to be considered. Two utilization levels were considered – 90% and 97%. The number of servers (operating rooms) was set to be 5 for the experiments. Using the average service time and these utilization values, two arrival rates were calculated. The first, corresponding to 90% utilization, was λ=2.07/hour. The second, corresponding to 97% utilization, was λ=2.283/hour

In summary, to model the operating room service node, 5 servers were modeled using 90% and 97% utilization. Two different distributions were modeled: Exponential (130) and Triangular (110, 137, and 147). Four different combinations were modeled, and within each of the combinations, the probability of a patient being classified low priority ranged from 0.1 to 0.9 in steps of 0.1. Therefore, a total of 36 combinations were tested. All combinations are shown below.

Table 3 – List of Combinations Tested

Utilization	**Distribution**	**P(Low Priority)**
90% Utilization	Exponential (130)	0.1
		0.2
		0.3
		0.4
		0.5
		0.6
		0.7
		0.8
		0.9
	Triangular (110, 137, 142)	0.1
		0.2
		0.3
		0.4
		0.5
		0.6
		0.7
		0.8
		0.9
97% Utilization	Exponential (130)	0.1
		0.2
		0.3
		0.4
		0.5
		0.6
		0.7
		0.8
		0.9
	Triangular (110, 137, 142)	0.1
		0.2
		0.3
		0.4
		0.5
		0.6
		0.7
		0.8
		0.9

The resulting average wait times from both the simulation and analytical models for each parameter combination are shown below. In the graphs, the x-axis shows the probability of an incoming patient being classified as low priority and the y-axis represents the average patient waiting time in minutes.

90% Utilization, Service Time Distribution Exponential with τ=130 minutes

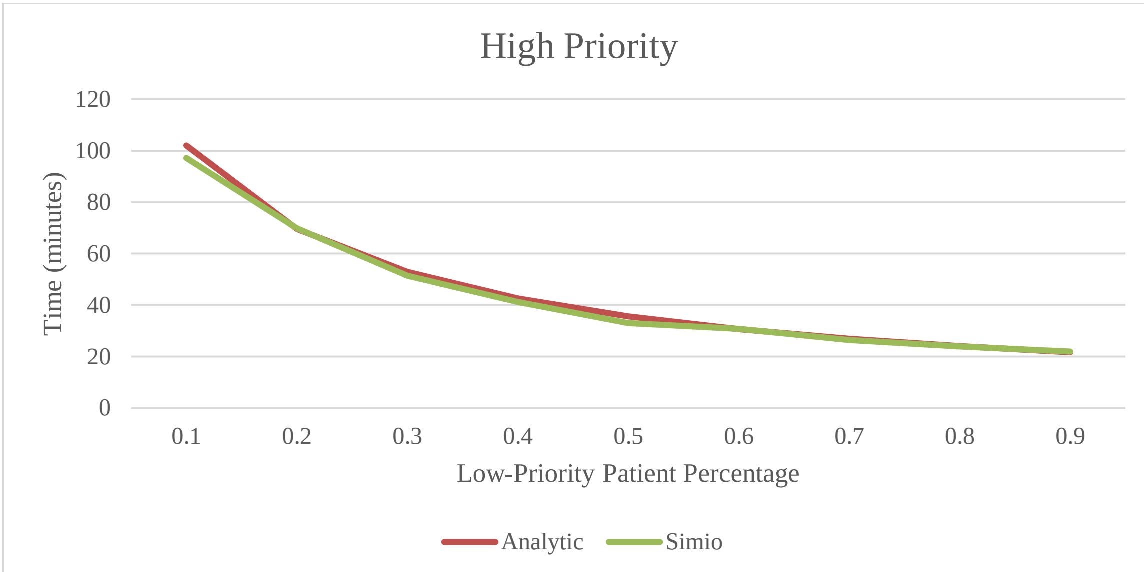

Figure 6 - High Priority Wait Time Results with 90% Utilization and Exponential Service Time Distribution

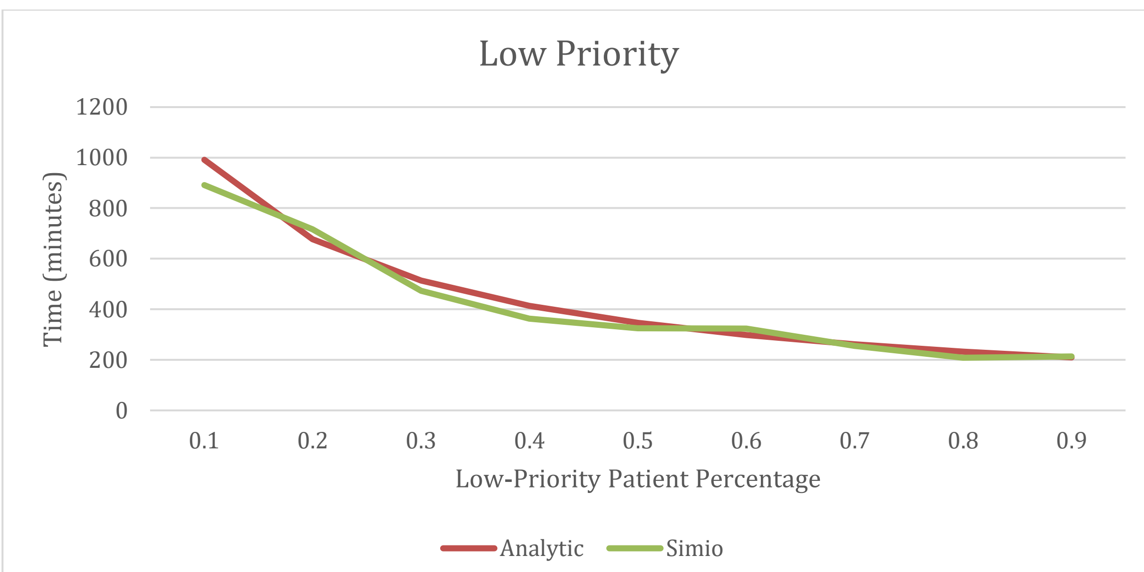

Figure 7 - Low Priority Wait Time Results with 90% Utilization and Exponential Service Time Distribution

90% Utilization, Service Time Distribution Triangular (110, 137, 142) minutes

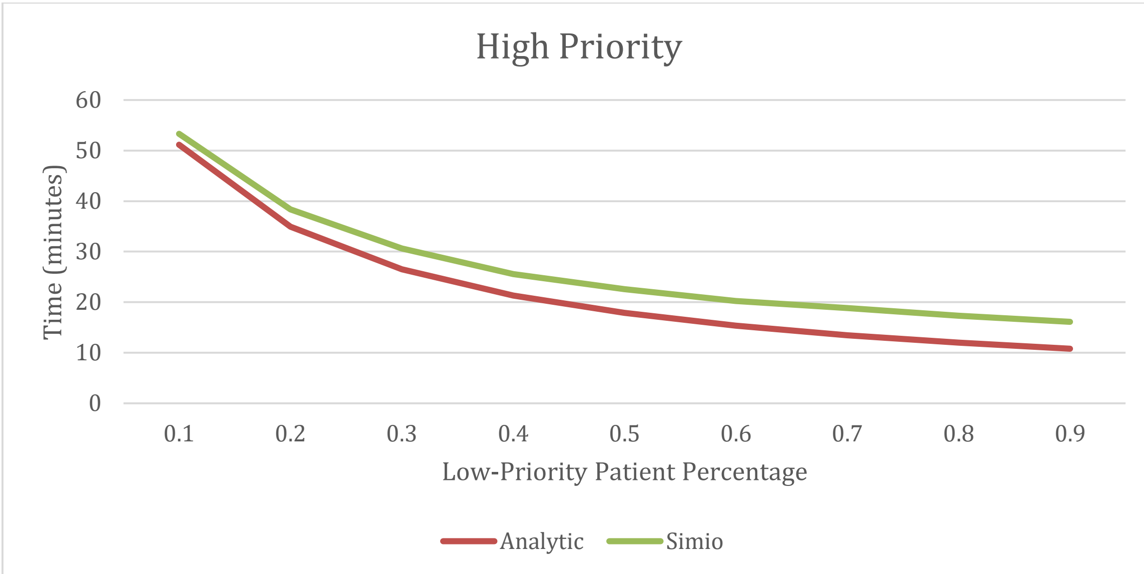

Figure 8 - High Priority Wait Time Results with 90% Utilization and Triangular Service Time Distribution

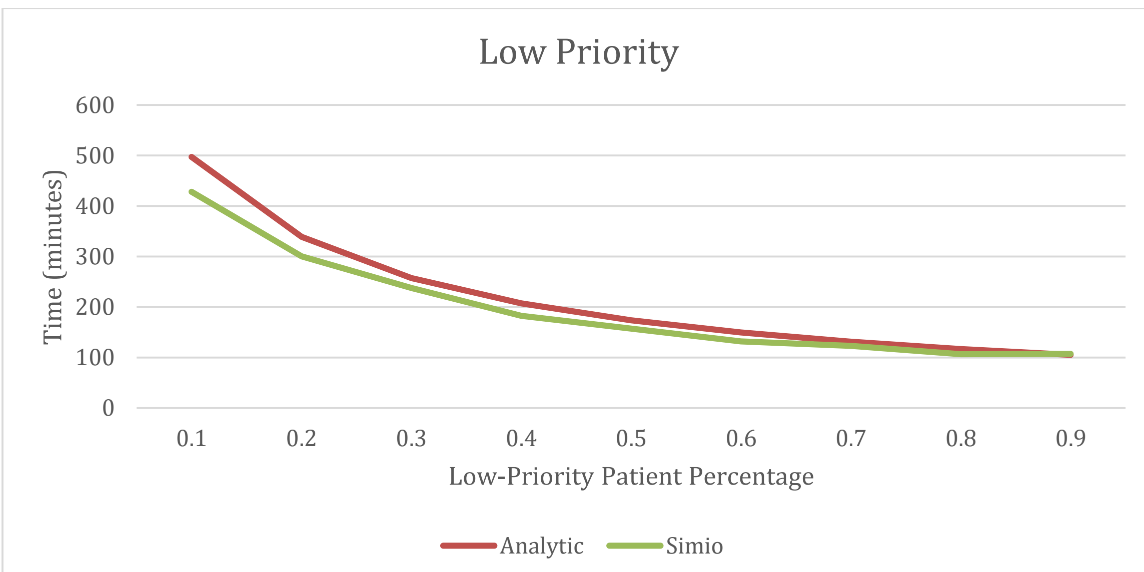

Figure 9 - Low Priority Wait Time Results with 90% Utilization and Triangular Service Time Distribution

97% Utilization, Service Time Distribution Exponential with τ=130 minutes

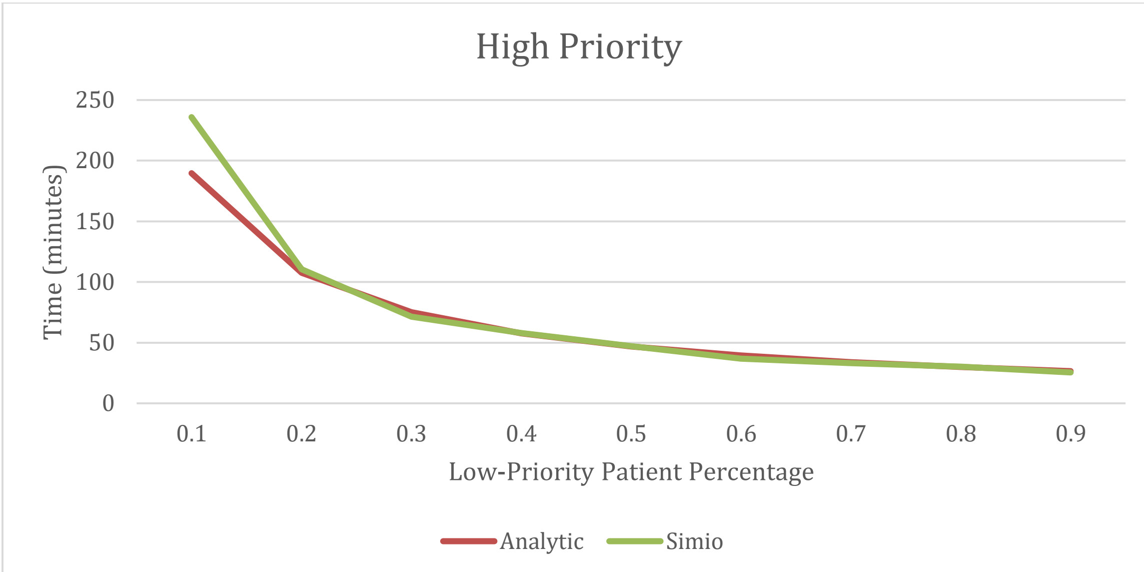

Figure 10 - High Priority Wait Time Results with 97% Utilization and Exponential Service Time Distribution

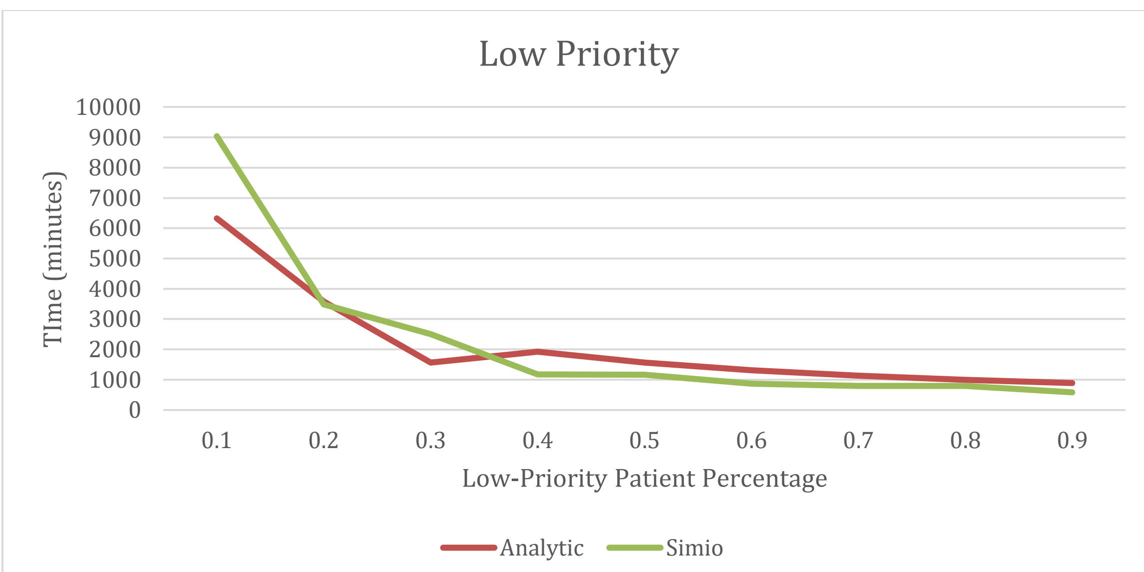

Figure 11 - Low Priority Wait Time Results with 97% Utilization and Exponential Service Time Distribution

97% Utilization, Service Time Distribution Triangular (110, 137, 142) minutes

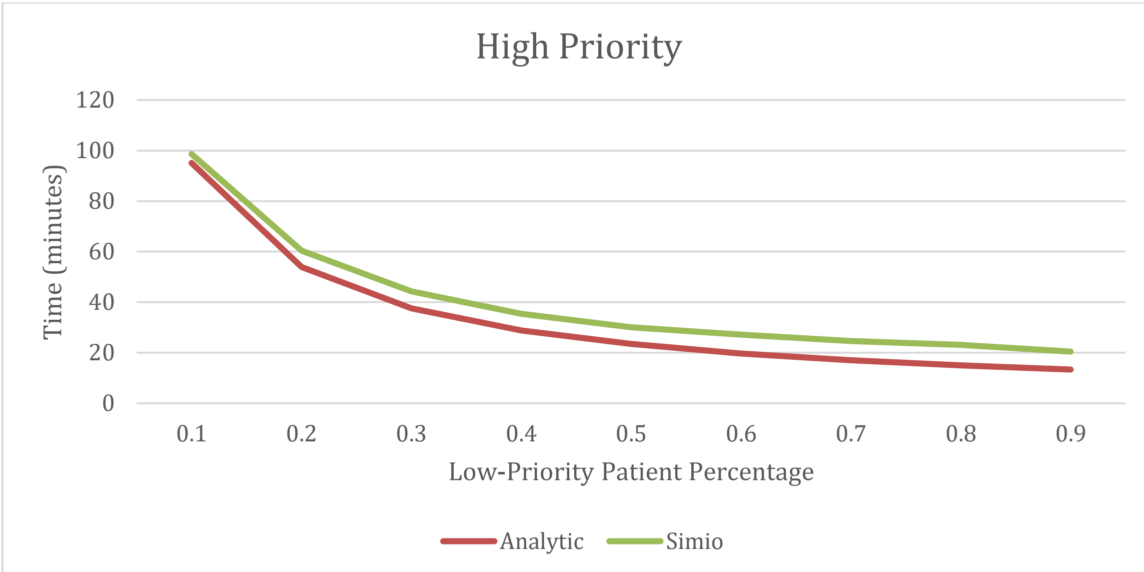

Figure 12 - High Priority Wait Time Results with 97% Utilization and Triangular Service Time Distribution

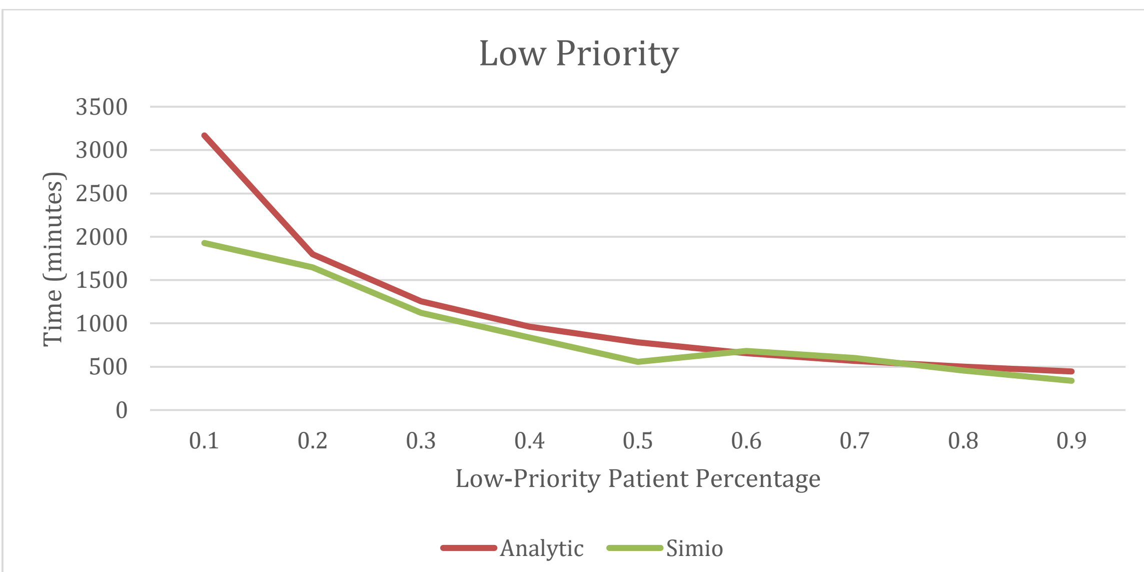

Figure 13 - Low Priority Wait Time Results with 97% Utilization and Triangular Service Time Distribution

Among all parameter combinations, the highest high priority wait time was 235.8 minutes. This occurred when utilization was 97%, the distribution was exponential, and the probability of an incoming patient being low priority was 10%. This was the only combination for which the high priority wait time was greater than two hours, which is the recommended maximum wait time for high priority patients as described by the triage guidelines. Additionally, research suggests that typically the percentage of low priority patients is around 80%-90% (*Emergency War Surgery*, 2004). For this range in the model, wait times are much lower than any other percentage of low priority patients.

Across all 72 parameter combinations, the analytic and simulation models were within 15% of each other 62.5% of the time. The difference between the two models was higher in some instances. This can likely be attributed to the high utilization that was used in these models. It should be noted that the analytical results for the exponential case are exact. In these cases, even the simulation results did not match with the exact results illustrating the difficulty is obtaining statistically accurate estimates from simulation models when the server utilization is close to 1. The agreement is much better at 90% utilization and worse at 97% utilization.

7.3 Treatment Areas Numerical Case

The first step in modeling the treatment areas was to determine the probability of exit from the Treatment Area. This value was calculated using the model for the open queuing network described in Chapter 6 with the average waiting time in queue from the priority queue model used as the total time in system. The search capability of Excel Solver was utilized to find the probability of exit.

After the probability of a patient exiting the system was estimated, the wait time values from the Check node and the total number of patients in the treatment area were calculated. For this numerical case, it was estimated that high priority patients must be checked every ten minutes and low priority patients must be checked every twenty minutes. The number of healthcare providers in each treatment area was set at 1. If wait times for the Check node were excessive, it could be concluded that the treatment area is understaffed and more healthcare workers are necessary to provide adequate patient care. The average number of patients in the treatment area is equivalent to the average number of beds needed. Results from this step are shown below for the 90% utilization case:

Table 4 - High Priority Results

High Priority						
Distribution	**p**	**P(exit)**	**Check Utilization**	**Total Time in Treatment Area (minutes)**	**Number in Treatment Area**	**Check Wait (minutes)**
Exponential	0.2	0.1747	31.61%	69.77	1.93	0.186
Triangular	0.2	0.3525	15.66%	34.39	0.95	0.123
Exponential	0.5	0.3387	10.19%	35.65	0.61	0.073
Triangular	0.5	0.6748	5.11%	17.86	0.31	0.053
Exponential	0.6	0.3920	7.04%	30.75	0.42	0.054
Triangular	0.6	0.7832	3.52%	15.37	0.21	0.038
Exponential	0.7	0.4477	4.62%	26.89	0.28	0.038
Triangular	0.7	0.8914	2.32%	13.49	0.14	0.025
Exponential	0.8	0.5020	2.75%	23.95	0.17	0.024
Triangular	0.8	1.0000	1.38%	12.02	0.08	0.015
Exponential	0.9	0.5563	1.24%	21.59	0.07	0.011

Table 5 - Low Priority Results

Low Priority						
Distribution	**p**	**P(exit)**	**Check Utilization**	**Total Time in Treatment Area (minutes)**	**Number in Treatment Area**	**Check Wait (minutes)**
Exponential	0.2	0.3270	42.22%	676.42	4.67	0.108
Triangular	0.2	0.6500	21.22%	339.19	2.34	0.056
Exponential	0.5	0.6420	53.72%	346.32	5.97	0.241
Triangular	0.5	0.1274	27.09%	173.66	3.00	0.119
Exponential	0.6	0.7480	55.34%	297.86	6.17	0.282
Triangular	0.6	0.1482	27.93%	149.36	3.09	0.139
Exponential	0.7	0.8540	56.54%	261.30	6.31	0.321
Triangular	0.7	0.1691	28.56%	131.03	3.16	0.157
Exponential	0.8	0.9610	57.46%	232.74	6.42	0.359
Triangular	0.8	0.1900	29.05%	116.71	3.22	0.175
Exponential	0.9	0.1068	58.17%	209.80	6.51	0.396

These results suggest that for the parameters tested, one healthcare worker per treatment area is likely sufficient.

For the case shown, very few beds are needed in the high priority treatment area, as the number in the treatment area is never above 2 for the parameters tested. However, more beds are needed in the low priority treatment area. This difference is due to the extended amount of time low priority patients wait for admission to the operating room when compared to the high priority patients.

CHAPTER VIII

CONCLUSION

8.1 Summary of Work

In this study, a literature review was completed that revealed the need for more research on modeling of triage systems after a mass casualty event. In response to this need, the goal of this study was to develop an analytical approach to modeling triage in order to determine the staffing and resource needs of hospitals after an MCI.

To complete this study, information was gathered over triage operations, such as the process that patients go through, staffing assignments, and typical times for operations, initial triage, and checks performed while patients wait for the operating room.

After relevant information was gathered, both analytical models and a simulation model in Simio were created to model triage operations and evaluate the level of care that patients receive under a variety of parameters, including staffing resources. These models were coded for two priorities, but could be easily modified to fit more than two priority categories.

8.2 Contributions

In preparedness for a mass casualty event, the ability of hospitals to make staffing and resource assignment decisions that are model-based is critically important. Understaffing could lead to patient conditions deteriorating as they wait for lifesaving operations.

This study details an approach of identifying the staff and resources needed through analytical models. Paired with a Decision Support System interface, the models developed in this study have the potential to aid hospital and emergency planners, giving them a detailed look at what patient wait times at different points in the system would look like under varying parameters. Using the models, this analysis can be accomplished without the need to draw from past data. This is especially significant because historical data isn't always available in crises that are unprecedented by any other time in history.

The study has also made a small contribution to the queueing body of knowledge by extending an exact model for a Markovian multi-server priority queueing model to develop approximations for the case with general arrival processes and service times. This simple extension greatly enhances the applicability of the model in practical situations with its ability to explicitly incorporate the effect of variability in arrival and service processes.

8.3 Future Work

Although all models are coded and validated, to make the approach outlined in this study more accessible to hospital planners and administrators that are responsible for developing emergency plans in the case of MCIs, a user interface will need to be developed. This user interface should have clear areas to input known parameters and test unknown parameters across a wide range of values so that users can draw from the range to gain additional insight.

After the user interface is developed, it will be prudent to evaluate its utility by soliciting feedback from individuals in hospital planning and administration roles. Ideally, they would be able to suggest areas that could be further developed that could benefit them and make the models more usable for those that could benefit the most.

By calculating the variance in waiting times and the number in the system, future work could focus on developing better estimates for the number of beds and other resources needed to provide care for patients with varying confidence levels.

Another area for future work is testing the models in a real-world scenario. It would be beneficial to work with a hospital to identify the input parameters that would be specific to an array of disasters. If an MCI occurred at a hospital that had used the models for planning, comparing the results from the models to real life data would be a valuable contribution to further validation of the models.

BIBLIOGRAPHY

About Disaster Management - IFRC. Retrieved from https://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management/

Abujudeh, H., Vuong, B., & Baker, S. (2005). Quality and operations of portable X-ray examination procedures in the emergency room: queuing theory at work. *Emergency Radiology, 11*(5), 262-266. doi:10.1007/s10140-005-0405-4

Adalja, A. A., Watson, M., Bouri, N., Minton, K., Morhard, R. C., & Toner, E. S. (2014). Absorbing citywide patient surge during Hurricane Sandy: a case study in accommodating multiple hospital evacuations. *Annals of Emergency Medicine, 64*(1), 66. doi:10.1016/j.annemergmed.2013.12.010

Bahadori, M., Mohammadnejhad, S., Ravangard, R., & Teymourzadeh, E. (2014). Using Queuing Theory and Simulation Model to Optimize Hospital Pharmacy Performance. *Iranian Red Crescent Medical Journal, 16*(3), 1-7. doi:10.5812/ircmj.16807

Bailey, N. T. J. (1952). A Study of Queues and Appointment Systems in Hospital Out-Patient Departments, with Special Reference to Waiting-Times. *Journal of the Royal Statistical Society. Series B (Methodological), vol. 14*, 185-199.

Bailey, N. T. J. (1954). On Queuing Processes with Bulk Service. *Journal of the Royal Statistical Society. Series B (Methodological), 16*, 80-87.

Betz, M., Stempien, J., Trivedi, S., & Bryce, R. (2017). A determination of emergency department pre-triage times in patients not arriving by ambulance compared to widely used guideline recommendations. *CJEM, 19*(4), 265-270. doi:10.1017/cem.2016.398

Bish, P. A., McCormick, M. A., & Otegbeye, M. (2016). Ready-JET-Go: Split Flow Accelerates ED Throughput. *Journal of Emergency Nursing, 42*(2), 114-119. doi:10.1016/j.jen.2015.06.003

Bonalumi, N. M., Bhattacharya, A., Edwards, C., Fasnacht, A., Mazzone, L., Stephens, K., . . . Swanson-Bierman, B. (2017). Impact of a Planned Workflow Change: Super Track Improves Quality and Service for Low-Acuity Patients at an Inner-City Hospital. *Journal of Emergency Nursing, 43*(2), 114-125. doi:10.1016/j.jen.2016.03.029

Busy ED keeps promise of 'door to doc' in 31 minutes. (2008). *Hospital case management : the monthly update on hospital-based care planning and critical paths, 16*(6), 87.

Cochran, J., & Bharti, A. (2006). Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science, 9*(1), 31-45. doi:10.1007/s10729-006-6278-6

Cochran, J., & Burdick, T. (2011). The Impact of the Door-To-Doc Emergency Department Patient Flow Model. *IIE Annual Conference. Proceedings*, 1-6.

Cochran, J. K., & Roche, K. T. (2009). A multi- class queuing network analysis methodology for improving hospital emergency department performance. *Computers and Operations Research, 36*(5), 1497-1512. doi:10.1016/j.cor.2008.02.004

Cohen, I., Mandelbaum, A., & Zychlinski, N. (2013). Minimizing Mortality in a Mass Casualty Event: Fluid Networks in Support of Modeling and Staffing. *IIE Transactions, 46*(7). doi:10.1080/0740817X.2013.855846

Connelly, L. G., & Bair, A. E. (2004). Discrete event simulation of emergency department activity: a platform for system-level operations research. *Acad Emerg Med, 11*(11), 1177-1185. doi:10.1197/j.aem.2004.08.021

Cotter, S. (2006, February 1, 2006). Treatment Area Considerations for Mass Casualty Incidents. *EMS World*.

de Bruin, A. M., Koole, G. M., & Visser, M. C. (2005). Bottleneck analysis of emergency cardiac in-patient flow in a university setting: an application of queueing theory. *Clinical and investigative medicine. Medecine clinique et experimentale, 28*(6), 316. doi:10.1016/S0379-4172(06)60012-7

Dorton, S. L., & Liu, D. (2016). Effects of Baggage Volume and Alarm Rate on Airport Security Screening Checkpoint Efficiency using Queuing Networks and Discrete Event Simulation. *Human Factors and Ergonomics in Manufacturing & Service Industries, 26*(1), 95-109. doi:10.1002/hfm.20616

Emergency War Surgery. (2004). Washington, D.C

Fawcett, W., & Oliveira, C. S. (2000). Casualty Treatment after Earthquake Disasters: Development of a Regional Simulation Model. *Disasters, 24*(3), 271-287. doi:10.1111/1467-7717.00148

Frederick S. Hillier, G. J. L. (2004). *Introduction to Operations Research*: McGraw-Hill Higher Education.

Gkritza, K., Niemeier, D., & Mannering, F. (2006). Airport security screening and changing passenger satisfaction: An exploratory assessment. *Journal of Air Transport Management, 12*(5), 213-219. doi:10.1016/j.jairtraman.2006.03.001

Gong, Q., & Batta, R. (2006). A Queue-Length Cutoff Model for a Preemptive Two-Priority $M/M/1$ System. *SIAM Journal on Applied Mathematics, 67*(1), 99. doi:10.1137/050648146

Hick, J. L., Barbera, J. A., & Kelen, G. D. (2009). Refining surge capacity: conventional, contingency, and crisis capacity. *Disaster medicine and public health preparedness, 3*(2 Suppl), S59. doi:10.1097/DMP.0b013e31819f1ae2

Houston, C., Sanchez, L. D., Fischer, C., Volz, K., & Wolfe, R. (2015). Waiting for Triage: Unmeasured Time in Patient Flow. *Western Journal of Emergency Medicine, 16*(1), 39-42. doi:10.5811/westjem.2014.11.22824

Huber-Wagner, S., Lefering, R., Kay, M. V., Stegmaier, J., Khalil, P. N., Paul, A. O., . . . Kanz, K. G. (2009). Duration and predictors of emergency surgical operations--basis for medical management of mass casualty incidents. *European journal of medical research, 14*, 532.

Kamath, M. (1994). RAQS.

Kennedy, K., Aghababian, R. V., Gans, L., & Lewis, C. P. (1996). Triage: Techniques and Applications in Decisionmaking. *Annals of Emergency Medicine, 28*(2), 136-144. doi:10.1016/S0196-0644(96)70053-7

Lodree, E. J., Altay, N., & Cook, R. A. (2017). Staff assignment policies for a mass casualty event queuing network. *Annals of Operations Research*. doi:10.1007/s10479-017-2635-8

Muhammet, G., & Ali Fuat, G. (2015). Simulation modelling of a patient surge in an emergency department under disaster conditions. *Croatian Operational Research Review, 6*(2), 429-443. doi:10.17535/crorr.2015.0033

Multivariable testing cuts door-to-doc times by 24%. (2007). *ED management : the monthly update on emergency department management, 19*(1), 6-7.

Organization, W. H. *Regional Training Course on Mass Casualty Management and Hospital Preparedness*. Retrieved from http://www2.wpro.who.int/internet/files/eha/dir/MCM training/MCM Module 4 Session 1 Hospital preparedness for mass casualty incidents.pdf

Ouyang, X., Patvivatsiri, L., & Montes, E. (2006). Application of Simulation Model in Healthcare System For Surge Capacity Evaluation. *IIE Annual Conference. Proceedings*, 1-6.

Phases of Disaster: Disaster Prepardness and Economic Recovery. Retrieved from http://restoreyoureconomy.org/disaster-overview/phases-of-disaster/

Pinto, L. R., De Campos, F. C. C., Perpetuo, I. H. O., & Ribeiro, Y. C. N. M. B. (2015). Analisys of hospital bed capacity via queuing theory and simulation. In (Vol. 2015-, pp. 1281-1292).

Sacco, W. J., Navin, D. M., Fiedler, K. E., Waddell Ii, R. K., Long, W. B., & Buckman, R. F. (2005). Precise Formulation and Evidence ‐ based Application of Resource ‐ constrained Triage. *Academic Emergency Medicine, 12*(8), 759-770. doi:10.1197/j.aem.2005.04.003

Slash door-to-doc time, boost patient approval. (2011). *Hospital case management : the monthly update on hospital-based care planning and critical paths, 19*(9), 142-143.

Song, C., & Zhuang, J. (2017). Two-stage security screening strategies in the face of strategic applicants, congestions and screening errors. *Annals of Operations Research, 258*(2), 237-262. doi:10.1007/s10479-015-2043-x

Sung, I., & Lee, T. (2016). Optimal allocation of emergency medical resources in a mass casualty incident: Patient prioritization by column generation. *European Journal of Operational Research, 252*(2), 623-634. doi:10.1016/j.ejor.2016.01.028

Takagi, H., Kanai, Y., & Misue, K. (2017). Queueing network model for obstetric patient flow in a hospital. *Health Care Management Science, 20*(3), 433-451. doi:10.1007/s10729-016-9363-5

TSA Boosts Efforts to Cut Waits at Airports -- WSJ. (2016). In. New York.

Ugarte, C. A., Ribka; Tieffenberg, Jacobo; Romig, Lou E.; Vu, Tien T. Planning and Triage in the Disaster Scenario. *American Academy of Pediatrics, Pediatric Education in Disasters Manual, Module 3*.

Whitt, W. (1983). The Queuing Network Analyzer. *The Bell System Technical Journal, 62*(9), 2779-2815.
Whitt, W. (1993). Approximations for the GI/G/m Queue. *Production and Operations Management, 2*, 114-161.

APPENDIX I

M/M/m Equations

$$P_0 = \left[\sum_{n=1}^{m-1} \frac{1}{n!} (\frac{\lambda}{\mu})^n + \frac{1}{m!} (\frac{\lambda}{\mu})^m (\frac{m\mu}{m\mu - \lambda}) \right]^{-1}$$

Estimated: Length of Queue:

$$L_q = \sum_{n=m}^{\infty} (n-m) P_n = \left[\frac{(\frac{\lambda}{\mu})^m \lambda\mu}{(m-1)!\,(m\mu - \lambda)^2} \right] P_O$$

Estimated Wait in Queue:

$$W_q = \frac{L_q}{\lambda}$$

VITA

Bailey Whitman Bonjour

Candidate for the Degree of

Master of Science

Thesis: MODELING AND ANALYSIS OF MASS CASUALTY TRIAGE SYSTEMS

Major Field: Industrial Engineering and Management

Biographical:

Education:

Completed the requirements for the Master of Science in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in May 2020.

Completed the requirements for the Bachelor of Science in Industrial Engineering and Management at Oklahoma State University, Stillwater, Oklahoma in 2017.

Professional Memberships: Institute of Industrial and Systems Engineers